Quadratic Programming

Quadratic programming is a special case of *non-linear* programming, and has many applications. One application is for optimal portfolio selection, which was developed by Markowitz in 1959 and won him the Nobel Prize in Economics.

1 Portfolio optimization, formulation

Before we present below an example of portfolio optimization, some basic knowledge of probability is discussed.

1.1 Digression to probability

The stock price in the future is a *random variable*, indicating that the price is random. Suppose we denote the stock price tomorrow (say) by S, then S is a random variable. Even though we cannot predict what exact value that S will be tomorrow, certain *distributional* property of S can be imposed to S (e.g. from historical data). For example, suppose today the stock price is \$40, and from historical data, you predict that tomorrow stock price with probability 1/4 will go up \$10, 1/2 stay the same, and 1/4 goes down \$10. In other words,

$$\mathbb{P}(S = 50) = \frac{1}{4}, \quad \mathbb{P}(S = 40) = \frac{1}{2}, \qquad \mathbb{P}(S = 30) = \frac{1}{4}.$$

What we have just written down is a *probability distribution*. It does not tell you exactly what tomorrow's stock price will be (nobody can tell you that), but gives you a very good idea what could possibly happen and what is the likelihood.

Given a random variable, and its distribution, one can compute the *expectation* (or, the average, the mean) of the random variable. The expectation of a random variable X is denoted by $\mathbb{E}X$. For example, we have

$$\mathbb{E}S = \frac{1}{4} \cdot 50 + \frac{1}{2} \cdot 40 + \frac{1}{4} \cdot 30 = 40.$$

Note that expectation is a *fixed* number, which is not random.

Lemma: For any random variables $\{X_1, X_2, \cdots, X_n\}$, and constant $\{a_1, a_2, \cdots, a_n\}$, we have

$$\mathbb{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1\mathbb{E}(X_1) + a_2\mathbb{E}(X_2) + \cdots + a_n\mathbb{E}(X_n). \tag{1}$$

The *variance* of a random variable X, denoted by $\mathrm{Var}(X)$, describes the variation of the random variable X. It is defined as

$$\mathrm{Var}(X) \doteq \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2. \tag{2}$$

Intuitively, if the deviation of X from $\mathbb{E}X$ tends to be larger with a bigger probability (or the variation of X is bigger), the variance tends to be bigger.

Lemma: $\mathrm{Var}(X) \geq 0$ for any random variable X, and the equality holds if and only if X is a constant.

The square root of the variance is called the *standard deviation*:

$$\sigma(X) \doteq \mathrm{Var}(X).$$

For example,

$$\mathrm{Var}(S) = \mathbb{E}(S - \mathbb{E}S)^2 = \mathbb{E}(S-40)^2 = \frac{1}{4}\cdot(50-40)^2 + \frac{1}{2}\cdot(40-40)^2 + \frac{1}{4}\cdot(30-40)^2 = 50,$$

and

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{50} = 5\sqrt{2}.$$

The generalization of variance is the so-called *covariance* of two random variables. Suppose X and Y are two random variables, define

$$\mathrm{Cov}(X,Y) \doteq \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right] = \mathbb{E}[XY] - \mathbb{E}X\cdot\mathbb{E}Y. \tag{3}$$

By definition it is clear that

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X), \qquad \mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X).$$

Remark: The covariance $\mathrm{Cov}(X,Y)$ describes the (linear) association between X and Y: if X tends to get bigger when Y gets bigger, then the covariance is positive; if X tends to get bigger when Y gets smaller, then the covariance is negative.

Lemma: We have the following formulae

$$\mathrm{Cov}(aX,bY) = ab\cdot\mathrm{Cov}(X,Y), \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y). \tag{4}$$

In particular, $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$. Here a, b are given constants.

1.2 Portfolio optimization

Consider an investor who has a fixed amount of money that can be invested in stocks and bonds. By manipulating the portfolio, the best he can do is to (1) maximize the expected return from his investment (2) minimize the risk associated with the portfolio (of course, there are many ways to define "risk", but here we define the risk as the variance of the return from the investment). However, the situation is that the two goals are contradictory: a portfolio that yields a large expected return is usually very risky, while less risky portfolio often gives a low return. Therefore, the approach one often adopts is to select a portfolio which attains an acceptable expected return while minimizing the variance. For example, an investor might seek a minimum variance portfolio with an expected return no less than 10%.

The formulation of this reduces to a quadratic programming (QP) problem. Consider the following example.

Example: Suppose an investor has 10 million dollars to invest in three stocks. Let S_j (random variable) denote the annual return of the j-th stock. For example, if $S_j = 0.1$, then \$1 invested in the j-th stock at the beginning of the year will be worth \$1.1 at the end of the year. We are given the following information:

$$\mathbb{E}S_1 = 0.14, \quad \mathbb{E}S_2 = 0.11, \quad \mathbb{E}S_3 = 0.10$$

and

$$\mathrm{Var}(S_1) = 0.20, \quad \mathrm{Var}(S_2) = 0.08, \quad \mathrm{Var}(S_3) = 0.18$$

and

$$\mathrm{Cov}(S_1, S_2) = 0.05, \quad \mathrm{Cov}(S_1, S_3) = 0.02, \quad \mathrm{Cov}(S_2, S_3) = 0.03.$$

The investor wants to construct a portfolio that attain an expected return no less than 12% and minimize the variance of the return of the portfolio.

Solution: Let x_j = amount (in millions) invested in the j-th stock. Then the annual return associated with this portfolio (a random variable) is

$$S = x_1 S_1 + x_2 S_2 + x_3 S_3.$$

The expected return of this portfolio is

$$\mathbb{E}(S) = x_1 \mathbb{E}S_1 + x_2 \mathbb{E}S_2 + x_3 \mathbb{E}S_3 = 0.14x_1 + 0.11x_2 + 0.10x_3.$$

The variance associated with the portfolio is

$$\begin{aligned} \mathrm{Var}(S) &= x_1^2 \mathrm{Var}(S_1) + x_2^2 \mathrm{Var}(S_2) + x_3^2 \mathrm{Var}(S_3) \\ &\qquad + 2x_1x_2\mathrm{Cov}(S_1, S_2) + 2x_1x_3\mathrm{Cov}(S_1, S_3) + 2x_2x_3\mathrm{Cov}(S_2, S_3) \\ &= 0.20x_1^2 + 0.08x_2^2 + 0.18x_3^2 + 0.10x_1x_2 + 0.04x_1x_3 + 0.06x_2x_3. \end{aligned}$$

The optimization problem is

$$\text{Minimize} \quad Z = 0.20x_1^2 + 0.08x_2^2 + 0.18x_3^2 + 0.10x_1x_2 + 0.04x_1x_3 + 0.06x_2x_3$$

such that

$$\begin{aligned} 0.14x_1 + 0.11x_2 + 0.10x_3 &\geq 0.12 \cdot 10 = 1.2 \\ x_1 + x_2 + x_3 &= 10. \end{aligned}$$

and $x_1, x_2, x_3 \geq 0$.

Remark: The variance can be written in a more compact form:

$$Z = x^T C x, \quad \text{with} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad C = \begin{bmatrix} 0.20 & 0.05 & 0.02 \\ 0.05 & 0.08 & 0.03 \\ 0.02 & 0.03 & 0.18 \end{bmatrix}.$$

Note here C is symmetric; i.e. $C^T = C$.

2 Wolfe's method to solve QP

Consider a general QP

$$\text{Minimize} \quad Z = \frac{1}{2}x^T Cx + p^T x$$

$$\text{such that} \quad Ax = b \text{ and } x \geq 0.$$

Here $A = (a_{ij})$ is an $m \times n$ matrix, $b = (b_i)$ is a $m \times 1$ vector, $x = (x_j)$ and $p = (p_j)$ are both $n \times 1$ vectors. The matrix C is an $n \times n$ matrix.

Assumption: The matrix C is symmetric ($C^T = C$) and is *positive definite*; i.e.

$$x^T Cx \geq 0 \text{ for every } n \times 1 \text{ vector } x.$$

Remark: For a portfolio optimization problem, this condition always holds (except for few pathological cases).

A result from linear algebra: A symmetric matrix C is positive definite if and only if there exists a matrix M such that $C = M^T M$.

Classical least square: The classical least-square is concerned with the same optimization problem but without constraints. And it is usually assumed that C is symmetric, positive-definite, and non-singular (invertible). The optimal solution can be obtained via taking derivatives and then setting the derivatives to zero. The optimal solution is

$$x^* = C^{-1}p.$$

Wolfe's method to solve QP is essentially a variant of simplex method for linear programming. The next result give the necessary and sufficient condition for a solution to be optimal for the QP.

Theorem: x^* is an optimal solution to the QP if and only if there exist an $m \times 1$ vector u^* and an $n \times 1$ vector v^* such that (x^*, u^*, v^*) solves the system of equations

$$\begin{aligned} Ax &= b \\ Cx + A^T u - v &= -p \\ x^T v &= 0. \end{aligned} \tag{5}$$

and $x \geq 0$, $v \geq 0$. (We do not require $u \geq 0$).

Proof: Write

$$q(x) \doteq \frac{1}{2}x^T Cx + p^T x.$$

"$\Rightarrow$": Suppose x^* is an optimal solution to the QP. Let y be an arbitrary vector such that

$$Ay = 0, \quad \text{and} \quad y_j \geq 0 \text{ if } x_j^* = 0 \quad (j = 1, \cdots, n). \tag{6}$$

It follows that $x^* + \varepsilon y$ is feasible for small enough $\varepsilon \geq 0$. By definition, we have

$$q(x^*) \quad \leq \quad q(x^* + \varepsilon y) = q(x^*) + \varepsilon \left(p + Cx^*\right)^T y + \frac{1}{2}\varepsilon^2 y^T C y.$$

Since this inequality holds for arbitrarily small ε, we must have

$$\left(p + Cx^*\right)^T y \geq 0. \tag{7}$$

In other words, for any vector y satisfies (6), the inequality (7) holds.

We denote by e_j the unit vector with j-th component 1 and other components zero. Then $y_j \geq 0$ is equivalent to $e_j^T y \geq 0$. Let $A_1, A_2, \cdots, A_m$ be the rows of matrix A. Then (6) says

$$\begin{aligned} A_i \cdot y &\geq 0, \quad i = 1, \cdots, m \\ -A_i \cdot y &\geq 0, \quad i = 1, \cdots, m \\ e_j^T \cdot y &\geq 0, \quad j \text{ such that } x_j^* = 0 \end{aligned}$$

By Farkas theorem, we have

$$p + Cx^* = \sum_{i=1}^{m} \rho_i A_i^T + \sum_{i=1}^{m} \sigma_i (-A_i)^T + \sum_j \tau_j e_j;$$

here the last summation is taken only for j such that $x_j^* = 0$.

We will now let $u_i^* \doteq -\rho_i + \sigma_i$, and set $u^* = (u_i^*)$. It follows that

$$\sum_{i=1}^{m} \rho_i A_i^T + \sum_{i=1}^{m} \sigma_i (-A_i)^T = -A^T u^*.$$

Also let $v^* = \sum_j \tau_j e_j$, then

$$Cx^* + A^T u^* - v^* = -p.$$

It remains to show that

$$(x^*)^T v^* = 0,$$

which holds since

$$(x^*)^T v^* = \sum_j \tau_j (x^*)^T e_j = \sum_j \tau_j x_j^* = 0.$$

"$\Leftarrow$": Suppose (x^*, u^*, v^*) is a solution to equations (5). We want to show that x^* is an optimal solution to the QP. Consider any feasible solution x, and write $y = x - x^*$. We have

$$q(x) = q(x^* + y) = q(x^*) + (p + Cx^*)^T y + \frac{1}{2} y^T C y.$$

By assumption,

$$(p + Cx^*)^T y = (v^* - A^T u^*)^T y = (v^*)^T y - (u^*)^T A y.$$

Since

$$Ay = Ax - Ax^* = b - b = 0,$$

and

$$(v^*)^T y = y^T v^* = x^T v^* - (x^*)^T v^* = x^T v^* \geq 0,$$

we have

$$q(x) - q(x^*) = (v^*)^T y + \frac{1}{2} y^T C y \geq \frac{1}{2} y^T C y \geq 0,$$

thanks to the positive definiteness of matrix C. We have $q(x^*) \leq q(x)$ for all feasible solution x, hence x^* is optimal. □

2.1 Wolfe's method

The preceding discussion says that solving QP is equivalent to solve the system of equations (5). The only "bad" equation is the nonlinear equation $x^T v = 0$. Without it, the equations can be written as

$$\begin{aligned} Ax &= b \\ Cx + A^T u - v &= -p \\ x \geq 0, \quad v \geq 0, &\qquad u \text{ free.} \end{aligned} \tag{8}$$

This is a system of linear equations, and we can use simplex algorithm to solve it.

Remark: How to formulate an LP to solve a general system of equations $Ax = b$, $x \geq 0$? The system of equations can also be written as

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

We can add artificial variables $d = (d_1, d_2, \cdots, d_m)$ formulate the LP

$$\text{Minimize} \quad Z = d_1 + d_2 + \cdots + d_m$$

such that

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n \pm d_1 &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \pm d_m &= b_m \end{aligned}$$

and $x \geq 0$, $d \geq 0$. Here for "$\pm d_i$" we mean "$+d_i$" if $b_i \geq 0$, and "$-d_i$" if $b_i < 0$.

An initial BFS is $x = 0$ and $d_i = |b_i|$. If the system equation $Ax = b, x \geq 0$ has a solution the optimal value Z^* will be zero; otherwise, the optimal value will be strictly positive.

Simplex algorithm will work for the equations (8). But we have to take into consideration the equation $x^T v = 0$. The trick here is that, the equation $x^T v = 0$ is going to be used as an *exclusion rule: x_j and v_j must not both be strictly positive for any $j = 1, \cdots, n$ at any stage of the simplex algorithm.* In other words, at any stage of the simplex algorithm for solving (8), if x_j is in the old basis ($x_j > 0$), we must not bring v_j into the basis; if v_j is in the old basis ($v_j > 0$), we must not bring x_j into the basis.

Example: Consider the QP

$$\text{Minimize} \quad Z = x_1^2 + x_2^2 + x_1 x_2 - 2x_1 - 3x_2$$

such that

$$x_1 + 2x_2 = 2$$

and $x \geq 0$.

Solution: For this problem,

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad p = \begin{bmatrix} -2 \\ -3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

It is equivalent to solve

$$\begin{aligned} x_1 + 2x_2 &= 2 \\ 2x_1 + x_2 + u - v_1 &= 2 \\ x_1 + 2x_2 + 2u - v_2 &= 3 \end{aligned}$$

such that $x \geq 0$, $v \geq 0$, u has no sign constraints, and $x^T v = 0$.

We write $u = u_1 - u_2$ with $u_1, u_2 \geq 0$ and add artificial variables d_1, d_2, d_3 as described in the previous Remark, we come up with the following initial tableau. Remember we want to maximize $Z = -d_1 - d_2 - d_3$

Basic Variable	Row	Z	x_1	x_2	u_1	u_2	v_1	v_2	d_1	d_2	d_3	RHS
Z	(0)	1	0	0	0	0	0	0	1	1	1	0
d_1	(1)	0	1	2	0	0	0	0	1	0	0	2
d_2	(2)	0	2	1	1	-1	-1	0	0	1	0	2
d_3	(3)	0	1	2	2	-2	0	-1	0	0	1	3

We have to make the coefficients of the basic variables 0 in Row (0) to start the simplex.

Basic Variable	Row	Z	x_1	x_2	u_1	u_2	v_1	v_2	d_1	d_2	d_3	RHS	Ratio
Z	(0)	1	-4	-5	-3	3	1	1	0	0	0	-7	
d_1	(1)	0	1	2*	0	0	0	0	1	0	0	2	$2/2 = 1 \leftarrow \min$
d_2	(2)	0	2	1	1	-1	-1	0	0	1	0	2	$2/1 = 2$
d_3	(3)	0	1	2	2	-2	0	-1	0	0	1	3	$3/2 = 1.5$

Here x_2 will enter the tableau. This is allowed since v_2 is non-basic. We arrive at

Basic Variable	Row	Z	x_1	x_2	u_1	u_2	v_1	v_2	d_1	d_2	d_3	RHS	Ratio
Z	(0)	1	-1.5	0	-3	3	1	1	2.5	0	0	-2	
x_2	(1)	0	0.5	1	0	0	0	0	0.5	0	0	1	
d_2	(2)	0	1.5	0	1	-1	-1	0	-0.5	1	0	1	$1/1 = 1$
d_3	(3)	0	0	0	$2*$	-2	0	-1	-1	0	1	1	$1/2 = 0.5 \leftarrow$min

Basic Variable	Row	Z	x_1	x_2	u_1	u_2	v_1	v_2	d_1	d_2	d_3	RHS	Ratio
Z	(0)	1	-1.5	0	0	0	1	-0.5	1	0	1.5	-0.5	
x_2	(1)	0	0.5	1	0	0	0	0	0.5	0	0	1	$1/0.5 = 2$
d_2	(2)	0	$1.5*$	0	0	0	-1	0.5	0	1	-0.5	0.5	$0.5/1.5 \leftarrow$min
u_1	(3)	0	0	0	1	-1	0	-0.5	-0.5	0	0.5	0.5	

Here x_1 will enter the tableau. This is allowed since v_1 is non-basic. We arrive at

Basic Variable	Row	Z	x_1	x_2	u_1	u_2	v_1	v_2	d_1	d_2	d_3	RHS	Ratio
Z	(0)	1	0	0	0	0	0	0	1	1	1	0	
x_2	(1)	0	0	1	0	0	$\frac{1}{3}$	$-\frac{1}{6}$	0.5	$-\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{6}$	
x_1	(2)	0	1	0	0	0	$-\frac{2}{3}$	$\frac{1}{3}$	0	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	
u_1	(3)	0	0	0	1	-1	0	-0.5	-0.5	0	0.5	0.5	

The optimal solution is therefore

$$(x_1^*, x_2^*) = (\frac{1}{3}, \frac{5}{6})$$